Novel Distance Measures for Vote Aggregation

Farzad Farnoud (Hassanzadeh), Behrouz Touri, and Olgica Milenkovic
University of Illinois, Urbana-Champaign, IL
E-mail: {hassanz1,touri1,milenkov}@illinois.edu

***Abstract*—We consider the problem of rank aggregation based on new distance measures derived through axiomatic approaches and based on score-based methods. In the first scenario, we derive novel distance measures that allow for discriminating between the ranking process of highest and lowest ranked elements in the list. These distance functions represent weighted versions of Kendall's τ measure and may be computed efficiently in polynomial time. Furthermore, we describe how such axiomatic approaches may be extended to the study of score-based aggregation and present the first analysis of distributed vote aggregation over networks.**

I. Introduction

Rank aggregation is a classical problem frequently encountered in social sciences, web search and Internet service analysis, expert opinion and voting theory [1]–[7]. The problem can be succinctly described as follows: a set of "voters" or "experts" is presented with a set of distinguishable entities (objects, individuals, movies), typically represented by the set $\{1, 2, \cdots, n\}$. The voters' task is to arrange the entities in decreasing order of preference and pass on their ordered lists to an aggregator. The aggregator outputs a single preference list used as a representative of all voters. Hence, one has to be able to adequately measure the quality of representation made by a vote aggregator. Two distinct analytical rank aggregation methods were proposed so far, namely, *distance-based* methods and *score-(position-)based* methods. In the first case, the quality of the aggregate is measured via a distance function that describes how close the aggregate is to each individual vote. In the second case, the aggregate is obtained by computing a score for each ranked entity and then arranging the entities based on their score. Well known distance measures include Kendall's τ and Spearman's Footrule [8].

The goal of this work is to propose two novel research directions in rank aggregation: one, which builds upon the existing work of distance-based aggregation, but expands the scope and applicability of vote-distances; and another, which sets the stage for analyzing score-based vote aggregations over networks. The results presented in the paper include a new set of voting-fairness axioms that lead to distance measures previously unknown in literature, as well as an analysis of consensus in distributed score-based voting systems.

Our work on aggregation distance analysis is motivated by the following observations: a) in many applications, the top of the ranking is more important than the bottom and so changes to the top of the list must result in a more significant change in the aggregate ranking than changes to the bottom of the list; b) ranked entities may have different degrees of similarity and often the goal is to find the most diverse, yet highest ranked entities. Hence, swapping elements that are similar should be penalized less than swapping those that are not. To the best of the authors' knowledge, the work of Sculley [5] represents the only method proposed so far for handling similarity in rank aggregation. Sculley presents an aggregation method, based on the use of Markov chains first introduced by Dwork et. al., with the goal of assigning similar ranks to similar items. A handful of results are known for rank aggregation distances that address the problem of positional relevance, i.e. the significance of the top versus the bottom of the ranking [7]. In this context, we introduce the notions of weighted Kendall distance and weighted Cayley distance, both capable of addressing the top versus bottom ranking issue, and provide axiomatic characterizations for these distance measures.

The work on vote aggregation over networks considers the issue of reaching consensus about the aggregate ranking in an arbitrary network, either through local interactions or based on a gossip algorithms. The assumption is that voters are connected through a social network that allows them to adjust their votes based on the opinions of their neighbors or randomly chosen network nodes, or even based on exogenous opinions. For a special type of score-based scheme – Borda's rule – we show that convergence to a vote consensus occurs and we determine the rate of convergence. The analysis of rank aggregation over networks for distance-based aggregation rules, and in particular for Kendall's τ and the weighted Kendall distance, is postponed to the full version of the paper.

The paper is organized as follows. An overview of relevant concepts, definitions, and terminology is presented in Section II. Weighted Kendall distance measures and extensions thereof, as well their axiomatic definitions, are presented in Sections III and IV. Section V is devoted to the analysis of gossip algorithms for rank aggregation.

II. Preliminaries

Suppose one is given a set $\Sigma = \{\sigma_1, \sigma_2, \cdots, \sigma_m\}$ of rankings, where each ranking σ represents a permutation in $\mathbb{S}_n$, the symmetric group of order n.

Given a distance function d over the permutations in $\mathbb{S}_n$, the distance-based aggregation problem can be stated as

$$\min_{\pi \in \mathbb{S}_n} \sum_{i=1}^{m} \mathsf{d}(\pi, \sigma_i).$$

In words, the goal is to find a ranking π with minimum cumulative distance from Σ. Clearly, the choice of the distance function d is an important feature for all distance-based rank

aggregation methods. Many distance measures in use were derived by starting from a reasonable set of axioms and then showing that the given distance measure is a unique solution under the given set of axioms[1]. A distance function derived in this manner is Kendall's τ distance, based on Kemeny's axioms [1].

On the other hand, score-based methods are centered around aggregators that assign scores to objects based on their positions in the rankings of Σ. Objects are then sorted according to their scores to obtain the aggregate ranking. One of the best known rules in this family is Borda's aggregation rule, introduced by Jean-Charles de Borda [10] wherein, for each ranking σ_i, object j receives score $b_i^j = \sigma_i^{-1}(j)$. The average score of object j is $\bar{b}^j = \frac{1}{m}\sum_{i=1}^{m} b_i^j$. The aggregate ranking is obtained by assigning the highest rank to the object with the lowest average score, the second highest rank to the object with the second lowest average score and so on. Borda's method also has an axiomatic underpinning: in the context of social choice functions, Young [11] presented a set of axioms that showed that Borda's rule is the unique social choice function that satisfies the given axioms. A social choice function is a rule indicating a set of winners when votes are given as rankings. Note that although similar, a social choice function differs from an aggregation rule; while a social choice function returns a set of winners, an aggregation rule ranks all objects.

In what follows, we introduce the notation used throughout the paper and provide a novel proof for the uniqueness of Kendall's τ distance function for a set of reduced Kemeny axioms [1].

Let $e = 12\cdots n$ denote the identity permutation (ranking).

Definition 1. A transposition of two elements $a, b \in [n]$ in a permutation π is the swap of elements in positions a and b, and is denoted by $(a\,b)$. In general, we reserve the notation τ for an arbitrary transposition and when there is no confusion, we consider a transposition to be a permutation itself. If $|a-b| = 1$, the transposition is referred to as an *adjacent transposition.*

It is well known that any permutation may be reduced to e via transpositions or adjacent transpositions. The former process is referred to as sorting, while the later is known as sorting with adjacent transpositions. The smallest number of adjacent transpositions needed to sort a permutation π is known as the inversion number of the permutation. Equivalently, the corresponding distance $\mathsf{d}(e,\pi)$ is known as the Kendall's τ distance. The Kendall's τ can be computed in time $O(n^2)$.

We also find the following set useful in our analysis,

$$A(\pi,\sigma) = \big\{(\tau_1,\cdots,\tau_m): \\ m\in\mathbb{N}, \sigma = \pi\tau_1\cdots\tau_m, \tau_i = (a_i\ a_i+1), i\in[m]\big\}$$

i.e., the set of all sequences of adjacent transpositions that transform π into σ.

[1]This is to be contrasted with the celebrated Arrow's impossibility theorem [9].

For a ranking $\pi \in \mathbb{S}_n$ and $a, b \in [n]$, π is said to *rank a before b* if $\pi^{-1}(a) < \pi^{-1}(b)$. We denote this relationship as $a <_\pi b$. Two rankings π and σ *agree* on a pair $\{a,b\}$ of elements if both rank a before b or both rank b before a. Furthermore, the two rankings π and σ *disagree* on the pair $\{a,b\}$ if one ranks a before b and the other ranks b before a.

For example, consider $\pi = 1234$ and $\sigma = 4213$. We have that $4 <_\sigma 1$ and that π and σ agree for $\{2,3\}$ but disagree for $\{1,2\}$.

Definition 2. A ranking ω is said to be *between* two rankings π and σ, denoted by π–ω–σ, if for each pair of elements $\{a,b\}$, ω either agrees with π or σ (or both). The rankings $\pi_0,\cdots,\pi_m$ are said to be on a line, denoted by π_0–π_1–$\cdots$–π_m, if for every i, j, and k for which $0 \le i < j < k \le m$, we have π_i–π_j–π_k.

The basis of our subsequent analysis is the following set of axioms required for a rank aggregation measure, first introduced by Kemeny [1]:

Axioms I

1) d is a metric.
2) d is left-invariant, i.e. $\mathsf{d}(\sigma\pi,\sigma\omega) = \mathsf{d}(\pi,\omega)$, for any $\pi,\sigma,\omega \in \mathbb{S}_n$. In words, relabeling of objects should not change the distance between permutations.
3) For any π,σ, and ω, $\mathsf{d}(\pi,\sigma) = \mathsf{d}(\pi,\omega) + \mathsf{d}(\omega,\sigma)$ if and only if ω is between π and σ. This axiom may be viewed through a geometric lens: the triangle inequality has to be satisfied for all points that lie on a "straight line" between π and σ.
4) The smallest positive distance is one. This axiom is only used for normalization purposes.

Kemeny's original exposition included a fifth axiom which we restate for completeness: If two rankings π and σ agree except for a segment of k elements, the position of the segment within the ranking is not important. Here, a segment represents a set of objects that are ranked consecutively - i.e., a substring of the permutation. As an example, this axiom implies that

$$\mathsf{d}(123\underbrace{456}, 123\underbrace{654}) = \mathsf{d}(1\underbrace{456}23, 1\underbrace{654}23)$$

where the segment is denoted by braces. This axiom is redundant since an equally strong statement follows from the other four axioms, as we demonstrate below. Our alternative proof of Kemeny's result also reveals a simple method for generalizing the axioms in order to arrive at weighted distance measures.

Lemma 3. *For any d that satisfies Axioms I, and for any set of permutations $\pi_0,\cdots,\pi_m$ such that π_0–π_1–$\cdots$–π_m, one has*

$$\mathsf{d}(\pi_0,\pi_m) = \sum_{k=1}^{m}\mathsf{d}(\pi_{k-1},\pi_k).$$

Proof: The lemma follows from Axiom I.3 by induction. ∎

Lemma 4. *For any* d *that satisfies Axioms I, we have that* $\mathsf{d}\left(\left(i\ i+1\right), e\right) = \mathsf{d}\left((12), e\right), i \in [n-1]$.

Proof: We show that $\mathsf{d}\left(\left(23\right), e\right) = \mathsf{d}\left(\left(12\right), e\right)$. Repeating the same argument used for proving this special case gives $\mathsf{d}\left(\left(i\ i+1\right), e\right) = \mathsf{d}\left(\left(i-1\ i\right), e\right) = \cdots = \mathsf{d}\left(\left(12\right), e\right)$.

To show that $\mathsf{d}\left(\left(23\right), e\right) = \mathsf{d}\left(\left(12\right), e\right)$, we evaluate $\mathsf{d}(\pi, e)$ in two ways, where we choose $\pi = 32145\cdots n$.

On the one hand, note that π–ω–η–e where $\omega = \pi(12) = 23145\cdots n$ and $\eta = \omega(23) = 21345\cdots n$. As a result,

$$\begin{aligned}\mathsf{d}(\pi, e) &\overset{(\mathsf{a})}{=} \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \eta) + \mathsf{d}(\eta, e)\\ &= \mathsf{d}(\omega^{-1}\pi, e) + \mathsf{d}(\eta^{-1}\omega, e) + \mathsf{d}(\eta, e)\\ &= \mathsf{d}((12), e) + \mathsf{d}((23), e) + \mathsf{d}((12), e) \qquad (1)\end{aligned}$$

where (a) follows from Lemma 3.

On the other hand, note that π–α–β–e where $\alpha = \pi(23) = 31245\cdots n$ and $\beta = \alpha(12) = 13245\cdots n$. For this case,

$$\begin{aligned}\mathsf{d}(\pi, e) &= \mathsf{d}(\pi, \alpha) + \mathsf{d}(\alpha, \beta) + \mathsf{d}(\beta, e)\\ &= \mathsf{d}(\alpha^{-1}\pi, e) + \mathsf{d}(\beta^{-1}\alpha, e) + \mathsf{d}(\beta, e)\\ &= \mathsf{d}((23), e) + \mathsf{d}((12), e) + \mathsf{d}((23), e). \qquad (2)\end{aligned}$$

Expressions (1) and (2) imply that $\mathsf{d}\left(\left(23\right), e\right) = \mathsf{d}\left(\left(12\right), e\right)$. ■

Lemma 5. *For any* d *that satisfies Axioms I,* $\mathsf{d}(\gamma, e)$ *equals the minimum number of adjacent transpositions required to transform* γ *into* e.

Proof: Let

$$L(\pi, \sigma) = \{(\tau_1, \cdots, \tau_m) \in A(\pi, \sigma) : \pi\text{–}\pi\tau_1\text{–}\pi\tau_1\tau_2\text{–}\cdots\text{–}\sigma\}$$

be the subset of $A(\pi, \sigma)$ consisting of sequences of transpositions that transform π to σ by passing through a line. Let m be the minimum number of adjacent transpositions that transform γ into e. Furthermore, let $(\tau_1, \tau_2, \cdots, \tau_m) \in A(\gamma, e)$ and define $\gamma_i = \gamma\tau_1\cdots\tau_i, i = 0, \cdots, m$, with $\gamma_0 = \gamma$ and $\gamma_m = e$.

First, we show that γ_0–γ_1–$\cdots$–γ_m, that is, $(\tau_1, \tau_2, \cdots, \tau_m) \in L(\gamma, e)$. Suppose this were not the case. Then, there would exist $i < j < k$ such that γ_i, γ_j, and γ_k are not on a line, and thus, there would exists a pair $\{r, s\}$ for which γ_j disagrees with both γ_i and γ_k. Hence, there would be two transpositions, $\tau_{i'}$ and $\tau_{j'}$, with $i < i' \leq j$ and $j < j' \leq k$ that swap r and s. We could in this case remove $\tau_{i'}$ and $\tau_{j'}$ from $(\tau_0, \cdots, \tau_m)$ to obtain $(\tau_0, \cdots, \tau_{i'-1}, \tau_{i'+1}, \cdots, \tau_{j'-1}, \tau_{j'+1}, \tau_m) \in A(\gamma, e)$ with length $m-2$. This contradicts the optimality of the choice of m. Hence, $(\tau_1, \tau_2, \cdots, \tau_m) \in L(\gamma, e)$. Then Lemma 3 implies that

$$\mathsf{d}(\gamma, e) = \sum_{i=1}^{m} \mathsf{d}(\tau_i, e). \qquad (3)$$

From (3), it is clear that the minimum positive distance from the identity is obtained by some adjacent transpositions. But, Lemma 4 states that all adjacent transpositions have the same distance from the identity. Hence, from Axiom I.4, we have $\mathsf{d}(\tau, e) = 1$ for all adjacent transposition τ. This observation completes the proof of the lemma, since it implies that

$$\mathsf{d}(\gamma, e) = \sum_{i=1}^{m} \mathsf{d}(\tau_i, e) = \sum_{i=1}^{m} 1 = m.$$

■

Lemma 6. *For any* d *that satisfies Axioms I, we have*

$$\mathsf{d}(\pi, \sigma) = \min\left\{m : (\tau_1, \cdots, \tau_m) \in A(\pi, \sigma)\right\}.$$

That is, $\mathsf{d}(\pi, \sigma)$ *equals the minimum number of adjacent transpositions required to transform* π *into* σ.

Proof: We have $(\tau_1, \cdots, \tau_m) \in A(\pi, \sigma)$ if and only if $(\tau_1, \cdots, \tau_m) \in A(\sigma^{-1}\pi, e)$. Furthermore, left-invariance of d implies that $\mathsf{d}(\pi, \sigma) = \mathsf{d}(\sigma^{-1}\pi, e)$. Hence,

$$\begin{aligned}\mathsf{d}(\pi, \sigma) &= \mathsf{d}(\sigma^{-1}\pi, e)\\ &= \min\left\{m : (\tau_1, \cdots, \tau_m) \in A(\sigma^{-1}\pi, e)\right\}\\ &= \min\left\{m : (\tau_1, \cdots, \tau_m) \in A(\pi, \sigma)\right\}\end{aligned}$$

where the second equality follows from Lemma 5. ■

Theorem 7. *The unique distance* d *that satisfies Axioms I is*

$$\mathsf{d}_\tau(\pi, \sigma) = \min\left\{m : (\tau_1, \cdots, \tau_m) \in A(\pi, \sigma)\right\}.$$

Proof: The fact that d_τ satisfies Axioms I can be easily verified. Uniqueness follows from Lemma 6. ■

III. Weighted Kendall Distance

Our proof of the uniqueness of Kendall's τ distance under Axioms I reveals an important insight: Kendall's measure arises due to the fact that adjacent transpositions have uniform costs, which is a consequence of the betweenness property of one of the axioms. If one had a ranking problem in which costs of transpositions either depended on the elements involved or their locations, the uniformity assumption had to be changed. As we show below, a way to achieve this goal is to redefine the axioms in terms of the betweenness property.

Axioms II

1) d is a pseudo-metric, i.e. a generalized metric in which two distinct points may be at zero distance.
2) d is left-invariant.
3) For any π, σ disagreeing for more than one pair of elements, there exists a ω such that $\mathsf{d}(\pi, \sigma) = \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \sigma)$.

Lemma 8. *For any distance* d *that satisfies Axioms II, and for distinct* π *and* σ*, we have*

$$\mathsf{d}(\pi, \sigma) = \min_{(\tau_0, \cdots, \tau_m) \in A(\pi, \sigma)} \sum_{i=1}^{m} \mathsf{d}(\tau_i, e).$$

Proof: First, suppose that π and σ disagree on one pair of elements. Then, we have $\sigma = \pi(a\ a+1)$ for some $a \in [n-1]$. For each $(\tau_0, \cdots, \tau_m) \in A(\pi, \sigma)$, there exists an index j such that $\tau_j = (a\ a+1)$ and thus

$$\sum_{i=1}^{m} \mathsf{d}(\tau_i, e) \geq \mathsf{d}(\tau_j, e) = \mathsf{d}(\left(a\ a+1\right), e)$$

implying

$$\min_{(\tau_0,\cdots,\tau_m)\in A(\pi,\sigma)} \sum_{i=1}^{m} \mathsf{d}(\tau_i, e) \geq \mathsf{d}((a\ a+1), e). \quad (4)$$

On the other hand, since $((a\ a+1)) \in A(\pi,\sigma)$,

$$\min_{(\tau_0,\cdots,\tau_m)\in A(\pi,\sigma)} \sum_{i=1}^{m} \mathsf{d}(\tau_i, e) \leq \mathsf{d}((a\ a+1), e). \quad (5)$$

From (4) and (5),

$$\min_{(\tau_0,\cdots,\tau_m)\in A(\pi,\sigma)} \sum_{i=1}^{m} \mathsf{d}(\tau_i, e) = \mathsf{d}((a\ a+1), e) = \mathsf{d}(\pi,\sigma)$$

where the last equality follows from the left-invariance of d.

Next, suppose π and σ disagree for more than one pair of elements. A sequential application of Axiom II.3 implies that

$$\mathsf{d}(\pi,\sigma) = \min_{(\tau_0,\cdots,\tau_m)\in A(\pi,\sigma)} \sum_{i=1}^{m} \mathsf{d}(\tau_i, e),$$

which proves the claimed result. ■

Definition 9. A distance d is termed a *weighted Kendall distance* if there is a nonnegative *weight function* φ over the set of adjacent transpositions such that

$$\mathsf{d}(\pi,\sigma) = \min_{(\tau_0,\cdots,\tau_m)\in A(\pi,\sigma)} \sum_{i=1}^{m} \varphi_{\tau_i}$$

where φ_τ is the weight assigned to transposition τ by φ.

Note that a weighted Kendall distance is completely determined by its weight function φ.

Theorem 10. *A distance* d *satisfies Axioms II if and only if it is a weighted Kendall distance.*

Proof: It follows immediately from Lemma 8 that a distance d satisfying Axioms II is a weighted Kendall distance by letting

$$\varphi_\tau = \mathsf{d}(\tau, e)$$

for every adjacent transposition τ.

The proof of the converse is omitted since it is easy to verify that a weighted Kendall distance satisfies Axioms II. ■

The weighted Kendall distance provides a natural solution for issues related to the importance of the top-ranked candidates. Due to space limitations, we refer the reader interested in other applications of weighted distances to our recent work [12].

Computing the Weighted Kendall Distance

Computing the weighted Kendall distance between two rankings for general weight functions is not a task as straightforward as computing the Kendall's τ distance. However, in what follows, we show that for an important class of weight functions – termed "monotonic" weight functions – the weighted Kendall distance can be computed efficiently.

Definition 11. A weight function $\varphi : \mathbb{A}_n \to \mathbb{R}^+$, where $\mathbb{A}_n$ is the set of adjacent transpositions in $\mathbb{S}_n$, is decreasing if $i > j$ implies that $\varphi_{(i\ i+1)} \leq \varphi_{(j\ j+1)}$. Increasing weight functions are defined similarly.

Decreasing weight functions are important as they can be used to model the significance of the top of the ranking by assigning higher weights to transpositions at the top of the list.

Suppose a transformation $\tau = (\tau_1, \cdots, \tau_m)$ of length m transforms π into σ. The transformation may be viewed as a sequence of moves of elements indexed by i, $i = 1, \ldots, m$, from position $\pi^{-1}(i)$ to position $\sigma^{-1}(i)$. Let the *walk* along which element i is moved by transformation τ be denoted by $p^{i,\tau} = \left(p_1^{i,\tau}, \cdots, p_{|p^{i,\tau}|+1}^{i,\tau}\right)$ where $\left|p^{i,\tau}\right|$ is the length of the walk $p^{i,\tau}$.

We investigate the lengths of the walks $p^{i,\tau}, i \in [n]$. Let $\mathcal{I}_i(\pi,\sigma)$ be the set consisting of elements $j \in [n]$ such that π and σ disagree on the pair $\{i, j\}$. Furthermore, let $I_i(\pi,\sigma) = |\mathcal{I}_i(\pi,\sigma)|$. In the transformation τ, all elements of $\mathcal{I}_i(\pi,\sigma)$ must be swapped with i by some $\tau_k, k \in [m]$. Each such swap contributes length one to the walk $p^{i,\tau}$ and thus, $\left|p^{i,\tau}\right| \geq I_i(\pi,\sigma)$.

It is easy to see that

$$\mathsf{d}_\varphi(\pi,\sigma) = \min_{\tau\in A(\pi,\sigma)} \sum_{i=1}^{n} \frac{1}{2} \sum_{j=1}^{\left|p^{i,\tau}\right|} \varphi_{\left(p_j^{i,\tau} p_{j+1}^{i,\tau}\right)}.$$

Considering individual walks, we may thus write

$$\mathsf{d}_\varphi(\pi,\sigma) \geq \sum_{i=1}^{n} \frac{1}{2} \min_{p^i\in P_i} \sum_{j=1}^{\left|p^i\right|} \varphi_{\left(p_j^i p_{j+1}^i\right)} \quad (6)$$

where, for each i, P_i is the set of all walks of length $I_i(\pi,\sigma)$ starting from $\pi^{-1}(i)$ and ending at $\sigma^{-1}(i)$. Since φ is decreasing, the minimum is attained by the walks $p^{i,\star} = (\pi^{-1}(i), \cdots, \ell_i - 1, \ell_i, \ell_i - 1, \cdots, \sigma^{-1}(i))$ where ℓ_i is the solution to the equation

$$\ell_i - \pi^{-1}(i) + \ell_i - \sigma^{-1}(i) = I_i(\pi,\sigma)$$

and thus $\ell_i = \left(\pi^{-1}(i) + \sigma^{-1}(i) + I_i(\pi,\sigma)\right)/2$.

We show next that there exists a transformation $\tau^\star$ such that $p^{i,\tau^\star} = p^{i,\star}$ and thus equality in (6) can be achieved. The transformation in question, $\tau^\star$, transforms π into σ in n rounds. In round i, $\tau^\star$ moves i through a sequence of adjacent transpositions from position $\pi^{-1}(i)$ to position $\sigma^{-1}(i)$. It can be seen that, for each i, $p^{i,\tau} = (\pi^{-1}(i), \cdots, \ell_i' - 1, \ell_i', \ell_i' - 1, \cdots, \sigma^{-1}(i))$ for some ℓ_i'. Since each transposition in τ decreases the number of inversions by one, ℓ_i' also satisfies the equation

$$\ell_i' - \pi^{-1}(i) + \ell_i' - \sigma^{-1}(i) = I_i(\pi,\sigma),$$

implying that $\ell_i' = \ell_i$ and thus $p^{i,\tau^\star} = p^{i,\star}$. Consequently, one has the following proposition.

Proposition 12. *For rankings* $\pi, \sigma \in \mathbb{S}_n$, *we have*

$$\mathsf{d}_\varphi(\pi,\sigma) = \sum_{i=1}^{n} \frac{1}{2} \left(\sum_{j=\pi^{-1}(i)}^{\ell_i - 1} \varphi_{(j\ j+1)} + \sum_{j=\sigma^{-1}(i)}^{\ell_i - 1} \varphi_{(j\ j+1)} \right)$$

where $\ell_i = \left(\pi^{-1}(i) + \sigma^{-1}(i) + I_i(\pi,\sigma)\right)/2$.

Example 13. Consider the rankings $\pi = 4312$ and $e = 1234$ and a decreasing weight function φ. We have $I_i(\pi, e) = 2$ for $i = 1, 2$ and $I_i(\pi, e) = 3$ for $i = 3, 4$. Furthermore,

$$\ell_1 = \frac{3+1+2}{2} = 3, \qquad p^{1,\star} = (3,2,1),$$
$$\ell_2 = \frac{4+2+2}{2} = 4, \qquad p^{2,\star} = (4,3,2),$$
$$\ell_3 = \frac{2+3+3}{2} = 4, \qquad p^{3,\star} = (2,3,4,3),$$
$$\ell_4 = \frac{1+4+3}{2} = 4, \qquad p^{4,\star} = (1,2,3,4).$$

The minimum weight transformation is

$$\tau^\star = \left(\underbrace{(32),(21)}_{1},\underbrace{(43),(32)}_{2},\underbrace{(43)}_{3}\right)$$

where the numbers under the braces are the element that is moved by the indicated transpositions. The distance between π and e is

$$\mathsf{d}_\varphi(\pi, e) = \varphi_{(12)} + 2\varphi_{(23)} + 2\varphi_{(34)}.$$

Note that the result above implies that at least for one class of interesting weight functions that capture the importance of the position in the ranking, the computation of the distance is of the same order of complexity as that of standard Kendall's τ distance. Hence, distance computation does not represent a bottleneck for the employment of weighted distance metrics.

IV. Generalizing Kemeny's Approach

We proceed by showing how Kemeny's axiomatic approach may be extended further to introduce a number of new distances metrics useful in different ranking scenarios.

The first distance applies when only certain subsets of transpositions are allowed – for example, when only elements of a class may be reordered to obtain an aggregated ranking.

Definition 14. Consider a subset $G = \{g_1, \cdots, g_m\}$ of $\mathbb{S}_n$ such that $g \in G$ implies that $g^{-1} \in G$. Rankings π and σ are $G-$adjacent if there exist $g \in G$ such that $\pi = \sigma g$.

A *$G-$transformation* of π into σ is a vector $(g_1, \cdots, g_k), k \in \mathbb{N}$, with $g_i \in G, i \in [k]$, such that $\sigma = \pi g_1 g_2 \cdots g_k$ where k is the length of the $G-$transformation. The set of $G-$transformations of π into σ is denoted by $A_G(\pi, \sigma)$. A *minimum $G-$transformation* is a $G-$transformation of minimum length.

Furthermore, ω is said to be $G-$between π and σ if there exists a minimal transformation $(g_1, \cdots, g_k)$ of π into σ such that $\omega = \sigma g_1 \cdots g_j$ for some $j \in [k]$.

Definition 15. For a subset G of $\mathbb{S}_n$, a function $\mathsf{d} : \mathbb{S}_n \to [0, \infty]$ is said to be a *uniform $G-$distance* if

1) d is a metric.
2) d is left-invariant.
3) For any $\pi, \sigma \in \mathbb{S}_n$, if ω is between π and σ, then $\mathsf{d}(\pi, \sigma) = \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \sigma)$.
4) The smallest positive distance is one.

Definition 3 also applies to $G-$betweenness and can be restated as follows.

Lemma 16. *For a uniform $G-$distance* d*, and for* $\pi_0, \cdots, \pi_m$ *such that* π_0–π_1–$\cdots$–π_m*, we have*

$$\mathsf{d}(\pi_0, \pi_m) = \sum_{k=1}^{m} \mathsf{d}(\pi_{k-1}, \pi_k).$$

Remark 17. For some choices of G, as in Lemma 4 and Lemma 20 in the next section, one may show that all elements of G have distance one from the identity. For such G, it is easy to see that the uniform $G-$distance d exists and is unique, with

$$\mathsf{d}(\pi, \sigma) = \min_m \{m : (\tau_1, \cdots, \tau_m) \in A_G(\pi, \sigma)\}.$$

Definition 18. For a subset G of $\mathbb{S}_n$, a function $\mathsf{d} : \mathbb{S}_n \to [0, \infty]$ is said to be a *weighted $G-$distance* if

1) d is a pseudo-metric.
2) d is left-invariant.
3) For any $\pi, \sigma \in \mathbb{S}_n$, if π and σ are not $G-$adjacent, there exists a ω between π and σ, distinct from both, such that $\mathsf{d}(\pi, \sigma) = \mathsf{d}(\pi, \omega) + \mathsf{d}(\omega, \sigma)$.

Remark 19. It is straightforward to see that the weighted $G-$distance d exists and is uniquely determined by the values $\mathsf{d}(g, e), g \in G$ as

$$\mathsf{d}(\pi, \sigma) = \min_{(\tau_1, \cdots, \tau_m) \in A_G(\pi, \sigma)} \sum_{i=1}^{m} \mathsf{d}(\tau_i, e)$$

where the minimum is taken over all $G-$transformations $(\tau_1, \cdots, \tau_m)$ of π into σ.

As an example, let G from Definitions 15 and 18 be the set

$$\mathbb{T}_n = \{(ab) : a, b \in [n], a \neq b\}$$

of all transpositions.

The following lemma states that for a uniform $\mathbb{T}_n-$distance, all transpositions have equal distance from identity.

Lemma 20. *For a uniform $\mathbb{T}_n-$distance* d*, we have*

$$\mathsf{d}((ab), e) = \mathsf{d}((cd), e)$$

for all transpositions (ab) *and* (cd).

Proof: For $\{a, b\} = \{c, d\}$, the lemma is obvious. We prove the lemma for the case that $a, b, c,$ and d are distinct. A similar argument applies when $\{a, b\}$ and $\{c, d\}$ have one element in common. The argument parallels that of Lemma 4.

Let $\pi = (abcd)$, $\omega = (ad)\pi$, $\eta = (cd)\omega$ and note that $e = (bc)\eta$. Since, π–ω–η–e by Lemma 16 and left-invariance of d, we have

$$\mathsf{d}(\pi, e) = \mathsf{d}((ad), e) + \mathsf{d}((cd), e) + \mathsf{d}((bc), e). \tag{7}$$

Similarly, let $\alpha = (bc)\pi$, $\beta = (ab)\alpha$, and note that $e = (ad)\beta$. This shows

$$\mathsf{d}(\pi, e) = \mathsf{d}((bc), e) + \mathsf{d}((ab), e) + \mathsf{d}((ad), e). \tag{8}$$

Equating the right-hand-sids of (7) and (8) yields $\mathsf{d}((ab), e) = \mathsf{d}((cd), e)$. ■

By combining Remark 17 and Lemma 20, we arrive at the following theorem.

Theorem 21. *The uniform $\mathbb{T}_n-$distance exists and is unique. Namely,*

$$\mathsf{d}(\pi, \sigma) = \min_m \{m : (\tau_1, \cdots, \tau_m) \in A_{\mathbb{T}_n}(\pi, \sigma)\},$$

(commonly known as Cayley's distance) is the unique $\mathbb{T}_n-$distance.

Furthermore, Remark 19 leads to the following theorem.

Theorem 22. *The weighted $\mathbb{T}_n-$distance d exists and is uniquely determined by the values $\mathsf{d}(\tau, e)$, $\tau \in \mathbb{T}_n$ as*

$$\mathsf{d}(\pi, \sigma) = \min_{(\tau_1, \cdots, \tau_m) \in A_{\mathbb{T}_n}(\pi, \sigma)} \sum_{i=1}^{m} \mathsf{d}(\tau_i, e).$$

The weighted transposition distance can be used to model similarities of objects in rankings wherein transposing two similar items induces a smaller distance than transposing two dissimilar items [12].

Remark 23. Note that the generalization of Kemeny's axioms may also be applied to arrive at a generalization of Borda's score-based rule. A step in this direction was proposed by Young [13], who showed that a set of axioms leads to a generalization of Borda's rule wherein the kth preference of each ranking receives a score s_k, not necessarily equal to k. This generalization of Borda's rule may also be used to address the problem of top versus bottom in rankings. In particular, one may assign Borda scores s_k to the kth preference with

$$s_k = \sum_{l=1}^{k-1} \phi_l,$$

where ϕ_k is decreasing in l. For example, swapping two elements at the top of the ranking of a given voter changes the scores of each of the two corresponding objects by ϕ_1 while a similar swap at the bottom of the ranking, changes the scores by ϕ_{n-1}. Since $\phi_1 \geq \phi_{n-1}$, changes to the top of the list, in general, have a more significant affect on the aggregate ranking.

V. Distributed Vote Aggregation

The novel distance metrics, scoring methods and underlying rank aggregation problems discussed in the previous sections may be viewed as instances rank aggregation of m agents over a fully connected graph: i.e. every agent has access to the ranking of all other agents and hence, fixing the aggregation distance or scores and aggregation method (Kendall, Borda,...) and assuming infinite computational power, each individual can find an aggregate ranking of the society. Thus, assuming the uniqueness of the aggregate ranking, agents come to a *consensus* over the aggregate ranking in one computational step. Nevertheless, one can consider the more general problem of reaching consensus about the aggregate ranking in an arbitrary network through local interactions. In this section, we consider this problem over general networks and provide an analysis of convergence for a specific choice of aggregation method: i.e. the Borda aggregation method. The analysis of aggregation methods for some other distance measures described in the paper is postponed to the full version of the paper.

Let $G = ([m], E)$ be a connected undirected graph over m vertices with the edge set E that represents the connectivity pattern of agents in a network[2]. As before, we assume that each agent $i \in [m]$ has a ranking σ_i over n entities. There are multiple ways of distributed aggregation of opinion in such a network, all of which are recursive schemes.

One way to perform distributed aggregation is through neighbor aggregation. In this method, at discrete-time instances $t = 0, 1, \ldots$, each agent maintains an estimate $\hat{\pi}_i(t)$ of the aggregate ranking. At time t, each agent exchange its believe with his neighboring agents. Then, at time $t+1$, agent i sets its believe $\hat{\pi}_i(t+1)$ to be the aggregate ranking of all the estimates of the neighboring agents at time t, including his own aggregation.

Another way to do distributed aggregation is through gossiping over networks [14]. Suppose that at each time instance we pick an edge $\{i, j\} \in E$ with probability $p_{ij} > 0$. Then, agents i and j exchange their estimates $\hat{\pi}_i(t)$ and $\hat{\pi}_j(t)$ at time t and they both let $\hat{\pi}_i(t+1) = \hat{\pi}_j(t+1)$ be the aggregation of $\hat{\pi}_i(t)$ and $\hat{\pi}_j(t)$.

A. Gossiping Borda Vectors

We describe next a distributed method using the Borda's scheme and gossiping over networks. Let $b_i = b_i(0)$ be the vector of the initial rankings of n entities for agent i (for Borda's method we have the specific choice of $b_i = \pi_i^{-1}$). The goal is to compute $\bar{b} = \frac{1}{m}\sum_{i=1}^{m} b_i(0)$. One immediate solution to find $\bar{b}$ is through gossiping over the network as described by the following algorithm:

Distributed Rank Aggregation:

1) At time $t \geq 0$, pick an edge $\{i, i'\} \in E$ with probability $P_{ii'} > 0$ where $\sum_{\{i,i'\} \in E} P_{ii'} = 1$,
2) Let i, i' exchange their estimate $b_i(t), b_{i'}(t)$ and let $b_i(t+1) = b_{i'}(t+1) = \frac{1}{2}(b_i(t) + b_{i'}(t))$,
3) For $\ell \neq i, i'$, let $b_\ell(t+1) = b_\ell(t)$.

As proven in [15], the above scheme approaches the average as t goes to infinity.

Lemma 24. *If $G = ([m], E)$ is connected, then we almost surely have $\lim_{t\to\infty} b_i(t) = \bar{b}$.*

Proof: The lemma is direct consequence of the results in [15]. ■

Note that in the distributed rank aggregation algorithm the ultimate goal is to find the correct ordering of $\bar{b} = \frac{1}{m}\sum_{i=1}^{m} b_i(0)$ rather than the vector $\bar{b}$ itself. Thus, it is not important that the estimates of the ranking vectors converges to $\bar{b}$, but that the estimates of the actual ranks are correct. In

[2] Many of the discussions in this section can be generalized for the case of time-varying networks

other words, if for some time t, for all agents i, the ordering of $b_i(t)$ matches the ordering of $\bar{b}$ for all $i \in [m]$, then the society has already achieved consensus over the *ranking* of the objects. Here, we derive a probabilistic bound on the number of iterations needed to probabilistically reach the optimum ranking.

Throughout the following discussions, without loss of generality we may assume that $\bar{b}$ is ordered[3], i.e. $\bar{b}^1 \leq \bar{b}^2 \leq \cdots \leq \bar{b}^n$. We say that t is a consensus time for the *aggregate ranking* if the ordering of $\bar{b}_i(t)$ matches the ordering of $\bar{b}$ for all $i \in [m]$. The following result follows immediately from this definition:

Lemma 25. *If t is a consensus time for the ranking, then any $t' > t$ is a consensus time for the ranking.*

Proof: It suffice to show the result for $t' = t + 1$. Let $\{i, i'\}$ be the edge that is chosen randomly at time t. Since t is a consensus time for the ranking, we have $b_i^1(t) \leq \cdots \leq b_i^n(t)$ and $b_{i'}^1(t) \leq \cdots \leq b_{i'}^n(t)$, and thus we also have

$$\begin{aligned} b_i^1(t+1) &= \frac{1}{2}\left(b_i^1(t) + b_{i'}^1(t)\right) \\ &\leq \cdots \leq b_i^n(t+1) \\ &= \frac{1}{2}\left(b_i^n(t) + b_{i'}^n(t)\right), \end{aligned}$$

which proves the claim. ■

Based on the lemma above, let us define *the consensus time* T for the ordering to be:

$$T = \min\{t \geq 0 \mid t \text{ is a consensus time for the ordering.}\}$$

Note that for the random gossip scheme, T is a random variable and if we have an adapted process for the random choice of edges, T is a stopping time. Our goal is to provide a probabilistic bound for T. For this, let $r^j = \min\left\{\bar{b}^{j+1} - \bar{b}^j, \bar{b}^j - \bar{b}^{j-1}\right\}$ and let $d^j = \max_i \bar{b}_i^j(0) - \min_i \bar{b}_i^j(0)$. That is, r_j is the minimum distance of the average rating of j from the neighboring objects and d_j is the spread of the initial ratings of the agents for the object j. Then, we have the following result.

Theorem 26. *For the consensus time T of the ordering we have*

$$P(T > t) \leq 4m\lambda_2^t(W) \sum_{j=1}^{n} \left(\frac{d^j}{r^j}\right)^2,$$

where $W = \sum_{\{i,i'\}\in E} P_{ii'}\left(I - \frac{1}{2}(e_i - e_{i'})(e_i - e_{i'})^T\right)$, $e_i = [0\ \cdots\ 0\ 1\ 0\ \cdots\ 0]^T$ *is an* $m \times 1$ *vector with* i*th element equal to one, and* $\lambda_2(W)$ *is the second largest eigenvalue of* W.

Proof: Let $b^j(t)$ be the vector obtained by the rating of the m agents at time t for object j and let $y^j(t) = b^j(t) - \bar{b}^j$. Note that if $\|y^j(t)\|^2 \leq \left(\frac{r^j}{2}\right)^2$, then this means that $|b_i^j(t) - \bar{b}^j| \leq \frac{r^j}{2}$ for all i. Thus, if for all $j \in [n]$ we have $\|y^j(t)\|^2 \leq \left(\frac{r^j}{2}\right)^2$, then it follows that:

$$b_i^j(t) \leq \bar{b}^j + \frac{r^j}{2} \leq \frac{1}{2}(\bar{b}^j + \bar{b}^{j+1}),$$

where the last inequality follows from the fact that $r^j \leq \bar{b}^{j+1} - \bar{b}^j$. Similarly, we have:

$$b_i^{j+1}(t) \geq \bar{b}^{j+1} - \frac{r^{j+1}}{2} \geq \frac{1}{2}(\bar{b}^{j+1} + \bar{b}^j),$$

which follows from $r^{j+1} \leq \bar{b}^{j+1} - \bar{b}^j$. Hence, we have

$$\begin{aligned} b_i^1(t) &\leq \frac{1}{2}(\bar{b}^2 + \bar{b}^1) \leq b_i^2(t) \leq \frac{1}{2}(\bar{b}^3 + \bar{b}^2) \\ &\leq \cdots \leq \frac{1}{2}(\bar{b}^{n-1} + \bar{b}^n) \leq b_i^m(t), \end{aligned}$$

and so t is a consensus time for the algorithm. Thus,

$$\{T > t\} \subseteq \bigcup_{j=1}^{n} \left\{\|y^j(t)\|^2 \geq \left(\frac{r^j}{2}\right)^2\right\}$$

and hence, using the union bound, we obtain

$$P(T > t) \leq \sum_{j=1}^{n} P\left(\|y^j(t)\|^2 \geq \left(\frac{r^j}{2}\right)^2\right). \tag{9}$$

Markov's inequality implies that

$$P\left(\|y^j(t)\|^2 \geq \left(\frac{r^j}{2}\right)^2\right) \leq \left(\frac{2}{r^j}\right)^2 E\left[\|y^j(t)\|^2\right].$$

Using the analysis in [15], it can be shown that

$$E\left[\|y^j(t)\|^2\right] \leq \lambda_2^t(W)\|y^j(0)\|^2 \leq m\left(d^j\right)^2.$$

Combining the above two relations, we find

$$P\left(\|y^j(t)\|^2 \geq \left(\frac{r^j}{2}\right)^2\right) \leq 4m\lambda_2^t \left(\frac{d^j}{r^j}\right)^2.$$

Replacing the last inequality in (9), proves the assertion. ■

Note that from [15], if G is connected, then we have $\lambda_2 < 1$ and thus the probability $P(T > t)$ decays exponentially.

[3]Throughout this section we use superscript to denote the ranking of objects.

References

[1] J. G. Kemeny, "Mathematics without numbers," *Daedalus*, vol. 88, no. 4, pp. pp. 577–591, 1959.

[2] W. D. Cook and M. Kress, "Ordinal ranking with intensity of preference," *Management Science*, vol. 31, pp. 26–32, 01 1985.

[3] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank aggregation revisited," *Manuscript (Available at: www.eecs.harvard.edu/ michaelm/CS222/rank2.pdf)*, 2001.

[4] C. Dwork, R. Kumar, M. Naor, and D. Sivakumar, "Rank aggregation methods for the web," in *Proceedings of the 10th international conference on World Wide Web*, pp. 613–622, ACM, 2001.

[5] D. Sculley, "Rank aggregation for similar items," in *Proceedings of the Seventh SIAM International Conference on Data Mining*, Citeseer, 2007.

[6] F. Schalekamp and A. van Zuylen, "Rank aggregation: Together we're strong," *Proc. of 11th ALENEX*, pp. 38–51, 2009.

[7] R. Kumar and S. Vassilvitskii, "Generalized distances between rankings," in *Proceedings of the 19th international conference on World wide web*, WWW '10, (New York, NY, USA), pp. 571–580, ACM, 2010.

[8] P. Diaconis, "Group representations in probability and statistics," *Lecture Notes-Monograph Series*, vol. 11, 1988.

[9] K. J. Arrow, *Social choice and individual values*. No. 12, Yale Univ Pr, 1963.
[10] J.-C. de Borda, "Mémoire sur les élections au scrutin," *Histoire de l'Académie royale des sciences*, 1784.
[11] H. P. Young, "An axiomatization of borda's rule," *Journal of Economic Theory*, vol. 9, no. 1, pp. 43 – 52, 1974.
[12] F. Farnoud and O. Milenkovic, "Sorting of permutations by cost-constrained transpositions," *Information Theory, IEEE Transactions on*, vol. 58, pp. 3 –23, Jan. 2012.
[13] H. Young, "Social choice scoring functions," *SIAM Journal on Applied Mathematics*, vol. 28, pp. 824–838, 06 1975.
[14] D. Shah, *Gossip Algorithms*, vol. 3. Foundations and Trends in Networking, 2009.
[15] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.